

Measurement entropy in Generalized Non-Signalling Theory cannot detect bipartite non-locality

Josh Cadney^{1,*} and Noah Linden¹

¹*Department of Mathematics, University of Bristol, Bristol BS8 1TW, U.K.*

(Dated: July 13, 2012)

We consider entropy in Generalized Non-Signalling Theory (also known as box world) where the most common definition of entropy is the measurement entropy. In this setting, we completely characterize the set of allowed entropies for a bipartite state. We find that the only inequalities amongst these entropies are subadditivity and non-negativity. What is surprising is that non-locality does not play a role - in fact any bipartite entropy vector can be achieved by separable states of the theory. This is in stark contrast to the case of the von Neumann entropy in quantum theory, where only entangled states satisfy $S(AB) < S(A)$.

I. INTRODUCTION

Entropy is a crucial concept in both classical and quantum information theory. The Shannon entropy was originally introduced as a measure of the uncertainty of a random variable [1], which turned out to have many applications in information theory, including optimal compression rates and channel capacities. Remarkably, the von Neumann entropy was introduced 20 years before the Shannon entropy, and was motivated by thermodynamic considerations [2]. It has found innumerable applications in quantum information theory, including its role as a measure of pure state entanglement [3, 4], and as the analogue of the Shannon entropy in many quantum coding theorems [5].

Given a multi-party quantum state ρ one can compute the von Neumann entropy of its various reduced states e.g. $S(A) := S(\rho_A)$, $S(AB) := S(\rho_{AB})$ etc., and so form the *entropy vector* of this state ρ . So for example, for two-party states, the entropy vector is $(S(A), S(B), S(AB))$. For N parties, the entropy vector lives in the vector space of $2^N - 1$ real dimensions. The question of which vectors can arise has been the subject of increasing interest recently, both in the quantum (von Neumann entropy) [6–8] and classical (Shannon entropy) [9–11] cases.

For example for two parties, both quantum entropies S_Q and classical entropies S_C are non-negative and satisfy *subadditivity*

$$\begin{aligned} S_Q(AB) &\leq S_Q(A) + S_Q(B), \\ S_C(AB) &\leq S_C(A) + S_C(B), \end{aligned} \tag{1}$$

However the space of achievable entropy vectors is different for the classical and quantum cases. Whereas classical entropies S_C satisfy *monotonicity*

$$S_C(AB) \geq S_C(A), \tag{2}$$

quantum entropies are more general and only satisfy the weaker Araki-Lieb inequality [12]

$$S_Q(AB) \geq S_Q(A) - S_Q(B). \tag{3}$$

In particular, a vector such as $(1, 1, 0)$ is achievable as a quantum entropy vector; it is the entropy vector of a singlet. However this vector is not achievable for any classical distribution; it does

*Electronic address: josh.cadney@bristol.ac.uk

not satisfy (2). Thus the space of entropy vectors seems to capture some of the differences between classical and quantum states. Indeed study of the space of achievable entropy vectors is a powerful tool in investigating multi-party entanglement of quantum states.

Mathematically, the space of entropy vectors is a cone [6], and characterizing this cone is an important problem in classical and quantum information theory [13]. The problem is completely solved for three or fewer parties in the classical and quantum cases [6, 14] (leading to different cones, of course); the cases of four or more parties (classical or quantum) remain open. One may understand a cone either by giving the inequalities or, dually, by the extremal rays. And, perhaps not surprisingly, points on these extremal rays typically correspond to interesting states. For example for quantum entropy vectors of two parties the extremal rays include $\lambda(1, 1, 0)$, $\lambda \geq 0$; a point on this ray may be achieved by the singlet, as mentioned above. For three parties one of the extremal quantum rays may be achieved by the GHZ state (see also [7, 15]).

With these observations in mind we turn now to so called “generalized probabilistic theories” (GPTs) [16]. These are theories which generalize classical and quantum theories, beginning from an operational viewpoint, where states are characterized by the output distributions of certain measurements. One aim of this field of research is to compare these more general theories with quantum theory, and in doing so gain some intuition as to ‘why’ Nature seems to prefer quantum theory.

Attempts have been made to introduce an entropy function within these general theories. The most popular seems to be the *measurement entropy* which satisfies many desirable properties for an entropy function [17]: it reduces to the Shannon and von Neumann entropies in classical and quantum theories respectively; it is always non-negative; and it is concave. In certain (quite broad) classes of theories, it is also subadditive and continuous.

Here we investigate features of the measurement entropy in ‘generalized non-signaling theory’ (GNST) [16] (also known as *box world* [18]) - the most famous and well studied GPT, which allows all non-local correlations that are non-signalling. Our first goal is to characterize the set of *entropy vectors*. It has already been noted that this entropy violates *strong subadditivity* and so the allowed entropy vectors are in some sense more general than the corresponding classical and quantum ones [17]. Our initial thought was that the space of achievable entropy vectors in GNST would reflect and shed light on the way this theory generalizes classical and quantum states.

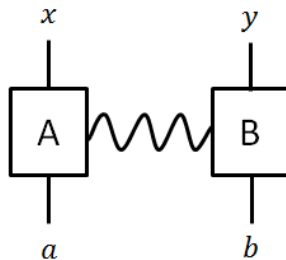
We are able to completely determine the set of bipartite GNST entropy vectors (up to the closure). We find this set to be the cone in \mathbb{R}^3 cut out by the non-negativity and subadditivity of the entropy *and no other inequalities*. This is in contrast with classical probability and quantum theory, where the analogous set is smaller due to the monotonicity (classical) and Araki-Lieb (quantum) inequalities.

What is very surprising, however, is that *every* entropy vector in GNST can be achieved by a separable state. This means that the measurement entropy is unable to detect non-locality. This is not true in quantum theory, where all separable states (but certainly not all states) satisfy the monotonicity relation $S(A) \leq S(AB)$ [19]; thus one may say that those quantum entropy vectors that do not satisfy monotonicity are the “truly” quantum ones.

The structure of the paper is as follows: in section 2 we briefly review GPTs; in section 3 we consider in some detail the allowed measurements in GNST; in section 4 we characterize the set of bipartite GNST entropy vectors; and in section 5 we consider the implications of this result. We close the paper with some concluding remarks.

II. GENERALIZED PROBABILISTIC THEORIES

It has been well known since Bell's theorem [20] that quantum theory admits correlations which are incompatible with any local classical theory. However, there are also correlations compatible with the no-signalling principle which cannot be produced by quantum theory. The most famous example is the PR-box [21].



Here, two parties (Alice and Bob) each own part of a bipartite system. Alice and Bob choose inputs $x, y \in \{0, 1\}$ respectively; the values of x and y correspond to different measurements on their systems. They then obtain outcomes $a, b \in \{0, 1\}$ according to the distribution:

$$p(a, b|x, y) = \begin{cases} \frac{1}{2} & \text{if } a \oplus b = xy \\ 0 & \text{else} \end{cases} \quad (4)$$

We know that quantum theory cannot produce such a distribution [22]. The motivation behind GPTs is to consider what kinds of physical theory could admit these and other general no-signalling correlations.

In order to form a new physical theory, we assume that the state of a system is determined by the outcome distributions of measurements on the system.

For an individual system, we assume that there is a set of k measurements, each with l outcomes, which determine the state uniquely. We call these the *fiducial* measurements. The state of the system is then a vector

$$\mathbf{p} = \begin{pmatrix} p(a = 0|x = 0) \\ p(a = 1|x = 0) \\ \vdots \\ p(a = l - 2|x = k - 1) \\ p(a = l - 1|x = k - 1) \end{pmatrix} \quad (5)$$

in a real vector space V . The values of k and l can vary from system to system (in the same way that different quantum systems have Hilbert spaces with different dimensions). For example, when $k = 1$ there is only one fiducial measurement, and we say that the system is *classical* since it is simply a classical random variable.

For a composite system, we make the further assumption that the fiducial measurements are those performed by simultaneously performing a fiducial measurement on each individual subsystem. This means that if the state is composed of n individual systems, then the state can be considered as a vector, \mathbf{p} , with components $p(a_1, \dots, a_n|x_1, \dots, x_n)$, also denoted $p(\mathbf{a}|\mathbf{x})$. Notice that \mathbf{p} naturally lives in the vector space $V_1 \otimes \dots \otimes V_n$, where V_i is the vector space containing the states of system i .

We now have many different types of system (each system contains some number, n , of individual systems, each of which has its own values for k and l). We obtain a physical theory

by specifying the sets of allowed states on each type of system. These sets must be convex to allow state mixing, and must satisfy the normalization condition:

$$\sum_{\mathbf{a}} p(\mathbf{a}|\mathbf{x}) = 1 \quad \forall \mathbf{x} \quad (6)$$

Further, all states must satisfy the no-signalling constraints:

$$\sum_{a_i} p(a_1, \dots, a_i, \dots, a_n | x_1, \dots, x_i, \dots, x_n) = \sum_{a_i} p(a_1, \dots, a_i, \dots, a_n | x_1, \dots, x'_i, \dots, x_n) \quad (7)$$

These constraints are important for two reasons: firstly, any state which violates these constraints would allow superluminal signalling. Secondly, they allow us to define the reduced state of a multipartite system:

$$p(\hat{\mathbf{a}}|\hat{\mathbf{x}}) := \sum_{a_i} p(\mathbf{a}|\mathbf{x}) \quad (8)$$

where e.g. $\hat{\mathbf{a}} = (a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_n)$, and we know that the sum does not depend on the value of x_i .

These theories can then be extended to include general measurements (this will be discussed in more detail in section III) and transformations, where we make further, physically motivated assumptions. For a full discussion see [16].

III. GENERALIZED NON-SIGNALLING THEORY

In this paper we consider a particular GPT known as ‘generalized non-signalling theory’ (GNST). GNST is the most general GPT in the sense that, for any type of system, the set of allowed states is *all* those which satisfy no-signalling. It is also known as ‘box world’ since we refer to individual systems in GNST as *boxes*. In this section we are especially interested in the measurements which the theory permits.

Measurements in GNST

Suppose we have a system with a set of allowed states \mathcal{S} . In a generalized probabilistic theory, an arbitrary measurement on the system (including, but not limited to, the fiducial measurements) has the following form: it is a set of pairs (r, μ_r) , where r is the outcome of the measurement, and μ_r is the corresponding *effect*. An effect is a linear map $\mu : \mathcal{S} \rightarrow [0, 1]$ (so that $\mu_r(\mathbf{p})$ is the probability that outcome r is obtained when the measurement is performed on state \mathbf{p}). To ensure that these probabilities always sum to 1, every measurement must have that $\sum_r \mu_r = u$, where u is the constant map $u(\mathbf{p}) = 1 \quad \forall \mathbf{p} \in \mathcal{S}$.

In GNST, any linear function $\mu : \mathcal{S} \rightarrow [0, 1]$ is an allowed effect, and any set of effects $\{\mu_r\}$ which sum to the unit map is an allowed measurement. Here we review what is known about the set of measurements in GNST, and prove a slight generalization of a result in [18] which we will use in Section IV.

Since effects are linear functionals, they must be of the form:

$$\mu_r(p) = \sum_{\mathbf{a}, \mathbf{x}} p(\mathbf{a}|\mathbf{x}) R_r(\mathbf{a}|\mathbf{x}) \quad (9)$$

for some vector \mathbf{R}_r (with entries indexed over \mathbf{a} and \mathbf{x}). We say that \mathbf{R}_r represents μ_r . However, note that there will be many vectors which represent each μ_r .

The following lemma is crucial.

Lemma 1 (Barrett [16, Appendix D]). *Every effect μ can be represented by some vector \mathbf{R} such that $0 \leq R(\mathbf{a}|\mathbf{x}) \leq 1$ for all \mathbf{a}, \mathbf{x} .*

Now suppose we have a composite system of many boxes. One way in which we can perform a measurement is to do the following:

- Choose one of the individual boxes and perform a fiducial measurement on that box.
- Based on the outcome of this measurement, choose another box and perform a fiducial measurement on that box.
- Repeat until all the boxes have been measured.

We call a measurement which has this form a *basic* measurement. The outcomes of a basic measurement are the outputs of the boxes, so that, in the language used above, the index r runs over all output vectors \mathbf{a} . The probability of obtaining output \mathbf{a} is equal to $p(\mathbf{a}|\mathbf{x}(\mathbf{a}))$, where $\mathbf{x}(\mathbf{a})$ is the list of inputs which are entered in the measurement when outputs \mathbf{a} are obtained. Thus, a vector representing the effect $\mu_{\mathbf{a}}$ is $\mathbf{R}_{\mathbf{a}}$ with components,

$$R_{\mathbf{a}}(\hat{\mathbf{a}}|\hat{\mathbf{x}}) = \begin{cases} 1 & \text{if } \hat{\mathbf{a}} = \mathbf{a} \text{ and } \hat{\mathbf{x}} = \mathbf{x}(\mathbf{a}) \\ 0 & \text{else} \end{cases} \quad (10)$$

In [18] it is shown that the only measurements which can be performed on systems of one or two boxes are basic measurements, or probabilistic mixtures of basic measurements. In section IV we will need the following generalization:

Lemma 2. *Let $A_1, A_2, B_1, \dots, B_n$ be a system of boxes, where only A_1 and A_2 are not classical. Then all measurements on this system are basic, or mixtures of basic measurements.*

The proof of this lemma is straightforward, and provided in the appendix.

Maximally informative measurements

A notion which will be important in the next section is that of a *maximally informative* or *fine-grained* measurement. Let $M = \{(r, \mu_r)\}$ and $N = \{(s, \nu_s)\}$ be two measurements, and denote their sets of possible outcomes by O_M and O_N respectively. We say that N is a *refinement* of M if O_N can be partitioned into sets P_r such that, for each r , $\mu_r = \sum_{s \in P_r} \nu_s$. In this case, N can be used to perform M (by performing N and returning r such that the outcome s is in the set P_r). The refinement is *trivial* if $\nu_s \propto \mu_r$ whenever $s \in P_r$. If M has no non-trivial refinement, then no other measurement reveals strictly more information about the state, and hence we call M maximally informative, or fine-grained.

The following lemma gives an important characterization of maximally informative measurements in GNST. Although the result may seem obvious in view of lemma 1, the proof requires a little effort and can be found in the appendix.

Lemma 3. *A GNST measurement $M = \{(r, \mu_r)\}$ is maximally informative if and only if every effect μ_r can be represented by a vector with only one non-zero entry, which is between 0 and 1.*

Remark. Lemma 3 ensures that basic measurements are maximally informative.

Suppose we have a composite system of subsystems X and Y (which may themselves be composite systems) and measurements M_X and M_Y on each system. One way to perform a measurement on the composite system would be to perform M_X on X and M_Y on Y independently. This is always an allowed measurement and we denote it $M_X \otimes M_Y$. Precisely, if M_X has effects μ_r represented by vectors \mathbf{R}_r and M_Y has effects ν_s represented by vectors \mathbf{R}_s , then $M_X \otimes M_Y$ has effects $\tau_{r,s}$ which can be represented by vectors $\mathbf{R}_r \otimes \mathbf{R}_s$. (The latter \otimes here is a genuine tensor product of vectors, as mentioned in section II).

Corollary 4. *If M_X is a maximally informative measurement on a box X and M_Y is maximally informative on Y , then $M_X \otimes M_Y$ is a maximally informative measurement on the composite system XY .*

Proof. If \mathbf{R}_X and \mathbf{R}_Y are vectors with one non-zero entry then so is $\mathbf{R}_X \otimes \mathbf{R}_Y$. Lemma 3 then implies the result. \square

IV. ENTROPY IN GNST

We are now in a position to introduce the entropy function which we will study in this paper. The measure of entropy we use is the measurement entropy, \hat{H} , which is defined as follows. Suppose we have a state \mathbf{p} on a system X . Then,

$$\hat{H}(X)_{\mathbf{p}} := \inf_{M \in \mathcal{M}} H_M(X)_{\mathbf{p}} \quad (11)$$

where $H_M(X)_{\mathbf{p}}$ is the Shannon entropy of the outcomes of measurement M on system X with state \mathbf{p} , and the infimum is taken over \mathcal{M} , the set of all maximally informative measurements. When it is clear which state we are referring to we will omit the subscript \mathbf{p} from the notation. The motivation for such a definition of entropy comes from the fact that in classical and quantum theories it is none other than the Shannon and von Neumann entropies respectively. It has previously been studied in [17, 23]. Lemma 3 implies that in GNST the infimum in (11) can be replaced by a minimum.

It is clear from the definition that the measurement entropy is always non-negative, since the Shannon entropy is non-negative. Before we can proceed with the main argument of the paper, we require two simple lemmas.

Lemma 5. *Measurement entropy in GNST is subadditive - for any state of a joint system XY we have that $\hat{H}(XY) \leq \hat{H}(X) + \hat{H}(Y)$.*

Proof. Suppose that M_X is the measurement on system X which achieves $\hat{H}(X)$, and M_Y is the measurement on system Y achieving $\hat{H}(Y)$. From section III we know that $M := M_X \otimes M_Y$ is a maximally informative measurement on system XY . Thus,

$$\hat{H}(XY) \leq H_M(XY) \quad (12)$$

$$\leq H_{M_X}(X) + H_{M_Y}(Y) \quad (13)$$

$$= \hat{H}(X) + \hat{H}(Y) \quad (14)$$

where the second inequality follows from the subadditivity of the Shannon entropy. \square

Remark. Note that this proof applies to any GPT in which the analogue of Corollary 4 holds. This argument was presented in [17].

Lemma 6. *If we restrict \mathcal{M} in (11) to include only basic measurements, then \hat{H} is additive on product states.*

Proof. Let $p(\mathbf{a}, \mathbf{b} | \mathbf{x}, \mathbf{y}) = p(\mathbf{a} | \mathbf{x})p(\mathbf{b} | \mathbf{y})$ be a product state of a joint system XY (where X is composed of n boxes, and Y is composed of m boxes). We proceed by induction on $n + m$.

Case $n + m = 1$. Wlog $n = 1, m = 0$. Then $\hat{H}(XY) = \hat{H}(X) = \hat{H}(X) + \hat{H}(Y)$.

Case $n + m \geq 2$. Let M be a measurement which achieves $\hat{H}(XY)$. Since M is a basic measurement it must begin by performing a fiducial measurement on one of the individual boxes. Wlog assume that M begins by performing measurement x_1 on box X_1 . Denote the Shannon entropy of the outcome of this measurement by $H_{x_1}(X_1)$. Now suppose that we have performed the measurement, and the result a_1 is known. Let \mathbf{q}_{a_1, x_1} be the remaining distribution on $X_2 \dots X_n Y$. Then from the rules of conditional probability:

$$q_{a_1, x_1}(a_2, \dots, a_n, \mathbf{b} | x_2, \dots, x_n, \mathbf{y}) = \frac{p(\mathbf{a}, \mathbf{b} | \mathbf{x}, \mathbf{y})}{p(a_1 | x_1)} \quad (15)$$

and notice that this is still a product state. Denote the remainder of the measurement, which has not yet been performed, by M_{a_1} . This is a basic measurement on $X_2 \dots X_n Y$. Then,

$$\hat{H}(XY)_{\mathbf{p}} = H_M(XY)_{\mathbf{p}} \quad (16)$$

$$= H_{x_1}(X_1) + \sum_{a_1} p(a_1 | x_1) H_{M_{a_1}}(X_2 \dots X_n Y)_{\mathbf{q}_{a_1, x_1}} \quad (17)$$

where the first line follows from the definition of M , and the second line from the grouping axiom of the Shannon entropy¹ [24]. From (17) we see that whenever a_1 can actually occur (i.e. $p(a_1 | x_1) > 0$), M_{a_1} must be a measurement which achieves $\hat{H}(X_2 \dots X_n Y)_{\mathbf{q}_{a_1, x_1}}$, else it would be possible to achieve a lower value for $\hat{H}(XY)_{\mathbf{p}}$. If a_1 cannot occur, we may just as well choose M_{a_1} to be such a measurement. Therefore we have,

$$\hat{H}(XY)_{\mathbf{p}} = H_{x_1}(X_1) + \sum_{a_1} p(a_1 | x_1) \hat{H}(X_2 \dots X_n Y)_{\mathbf{q}_{a_1, x_1}} \quad (18)$$

$$= H_{x_1}(X_1) + \sum_{a_1} p(a_1 | x_1) \hat{H}(X_2 \dots X_n)_{\mathbf{q}_{a_1, x_1}} + \hat{H}(Y)_{\mathbf{p}} \quad (19)$$

where the second line uses the induction hypothesis, and the fact that the reduced state of \mathbf{p} on system Y is the same as that of \mathbf{q}_{a_1, x_1} . By the grouping axiom, we see that the sum of the first two terms on the right hand side of (19) is the Shannon entropy of the outcomes of a measurement on system X . Therefore, by the definition of \hat{H} it follows that

$$\hat{H}(XY)_{\mathbf{p}} \geq \hat{H}(X)_{\mathbf{p}} + \hat{H}(Y)_{\mathbf{p}} \quad (20)$$

Since the proof of lemma 5 works also under the restriction to basic measurements, we arrive at the result. \square

Our aim is to investigate the set of GNST entropy vectors. Let us focus on the two party case. We know that a two party entropy vector is a vector in \mathbb{R}^3 , (x, y, z) , such that $x, y, z \geq 0$ and $z \leq x + y$. Are there any further constraints on the values of x, y, z ? We will show that in fact these conditions are all.

To this end, let \mathcal{C} be the set of points given by our necessary conditions:

$$\mathcal{C} := \{(x, y, z) \in \mathbb{R}^3 | x, y, z \geq 0, z \leq x + y\}. \quad (21)$$

¹ Suppose that we partition the outcomes of a measurement into groups labelled a_1, \dots, a_k and break up the measurement into two stages. First observe variable A - which group the outcome is in - and second observe variable B - the outcome from among that group. The grouping axiom states that the entropy of the overall measurement is equal to $H(A) + \sum_i p(a_i)H(B|A = a_i)$.

Then \mathcal{C} is a closed, convex cone (i.e. $v \in \mathcal{C}$ implies $\lambda v \in \mathcal{C}$ for all $\lambda \in \mathbb{R}_{\geq 0}$, and whenever $v_1, v_2 \in \mathcal{C}$, $v_1 + v_2 \in \mathcal{C}$ also). This means that we can characterize \mathcal{C} either by the linear inequalities which bound \mathcal{C} , or equivalently by its extremal rays. These extremal rays are the vectors:

$$e_1 := (1, 0, 1) \quad (22)$$

$$e_2 := (0, 1, 1) \quad (23)$$

$$e_3 := (1, 0, 0) \quad (24)$$

$$e_4 := (0, 1, 0). \quad (25)$$

Any $v \in \mathcal{C}$ can be written as $v = \lambda_1 e_1 + \lambda_2 e_2 + \lambda_3 e_3 + \lambda_4 e_4$, with $\lambda_i \geq 0$ for all i .

Consider the following joint probability distribution:

$$p(a, b) = \begin{cases} \frac{1}{2} & \text{if } b = 0 \\ 0 & \text{else} \end{cases} \quad (26)$$

where a, b both take values in $\{0, 1\}$. Then we can consider $p(a, b)$ to be the state of a two box system, in which each box has only one possible input. The entropies are then just the Shannon entropies of the different reduced states. Hence the entropy vector is $(1, 0, 1) = e_1$. We can similarly find a probability distribution achieving e_2 , and it is not hard to generalize these to distributions achieving λe_1 and λe_2 for any $\lambda \geq 0$. (Indeed, consider the distribution of two random variables - one of which is deterministic, and the other with Shannon entropy λ).

Now consider a system of two boxes, X, Y , where X has two possible inputs (0 and 1) and $N + 1$ outputs $(0, 1, \dots, N)$, and Y is a random variable (i.e. only one input) with two possible outputs (0 and 1). The distribution $p(a, b|x)$ is as follows:

$$p(a, b|x = 0) = \begin{cases} \frac{1}{2} & \text{if } a = b = 0 \\ \frac{1}{2N} & \text{if } a \geq 1, b = 1 \\ 0 & \text{else} \end{cases} \quad (27)$$

$$p(a, b|x = 1) = \begin{cases} \frac{1}{2} & \text{if } a = 0, b = 1 \\ \frac{1}{2N} & \text{if } a \geq 1, b = 0 \\ 0 & \text{else} \end{cases} \quad (28)$$

Since we have only two boxes, we know that the only allowed measurements are the basic measurements (or probabilistic mixtures of basic measurements, but a mixed measurement would not be optimal for achieving the minimum in (11)). This makes it easy to calculate the entropies. The distribution of X alone, for either input, is $(\frac{1}{2}, \frac{1}{2N}, \dots, \frac{1}{2N})$ and hence $\hat{H}(X) = 1 + \frac{1}{2} \log N$. The reduced distribution of Y is $(\frac{1}{2}, \frac{1}{2})$ and so $\hat{H}(Y) = 1$. Now consider the following measurement. First observe Y to obtain output b . If $b = 0$ set $x = 0$, otherwise set $x = 1$. Now observe X . With certainty we will find that $a = 0$, and the distribution of the measurement outcomes is $(\frac{1}{2}, \frac{1}{2})$. This implies that $\hat{H}(XY) \leq 1$, and in fact this measurement is optimal, i.e. $\hat{H}(XY) = 1$.

We have discovered that for every N , $(1 + \frac{1}{2} \log N, 1, 1)$ is an entropy vector.

We now alter this scenario by adding an extra character, ∞ , to the output alphabet of both boxes. Now for each $N \in \mathbb{N}$ consider the following distribution:

$$p(a, b|x = 0) = \begin{cases} \frac{\lambda_N}{2} & \text{if } a = b = 0 \\ \frac{\lambda_N}{2N} & \text{if } a \geq 1, b = 1 \\ 1 - \lambda_N & \text{if } a = b = \infty \\ 0 & \text{else} \end{cases} \quad (29)$$

$$p(a, b|x = 1) = \begin{cases} \frac{\lambda_N}{2} & \text{if } a = 0, b = 1 \\ \frac{\lambda_N}{2N} & \text{if } a \geq 1, b = 0 \\ 1 - \lambda_N & \text{if } a = b = \infty \\ 0 & \text{else} \end{cases} \quad (30)$$

where λ_N are (as yet unspecified) constants (between 0 and 1). Fix a positive real value, k , and set $\lambda_N = \frac{2k}{\log N}$. By the same reasoning as above, this distribution has entropy vector $(\lambda_N + \frac{\lambda_N}{2} \log N + h(\lambda_N), \lambda_N + h(\lambda_N), \lambda_N + h(\lambda_N)) = (\lambda_N + k + h(\lambda_N), \lambda_N + h(\lambda_N), \lambda_N + h(\lambda_N))$, here $h(q) := -q \log q - (1 - q) \log(1 - q)$. As $N \rightarrow \infty$, $\lambda_N \rightarrow 0$ and $h(\lambda_N) \rightarrow 0$ and so we have found entropy vectors arbitrarily close to $(k, 0, 0)$.

Theorem 7. *Every vector in \mathcal{C} is in the closure of the set of entropy vectors.*

Proof. Consider an arbitrary vector in \mathcal{C} , $v = \lambda_1 e_1 + \lambda_2 e_2 + \lambda_3 e_3 + \lambda_4 e_4$. We have found states $\rho_1, \rho_2, \rho_3, \rho_4$ whose entropy vectors are (arbitrarily close to) $\lambda_1 e_1, \lambda_2 e_2, \lambda_3 e_3, \lambda_4 e_4$ respectively, and ρ_1, ρ_2 are entirely classical, whereas ρ_3, ρ_4 have only one non-classical box. If we take $\sigma = \rho_1 \otimes \rho_2 \otimes \rho_3 \otimes \rho_4$, then σ has only two non-classical boxes. By lemma 2 the only measurements on σ are basic measurements, hence lemma 6 tells us that the measurement entropy is additive on product states here. Consequently, the entropy vector of σ is (arbitrarily close to) v . \square

V. RELATION TO NON-LOCALITY

In the previous section we gave the proof of the main technical result of the paper, but we have not yet delivered the punch line. The alert reader would have noticed that all the states used in the proof of theorem 7 are separable GNST states (and hence local in that they admit a local hidden variable description). For example, the state given by (27) and (28) can be decomposed in the following way:

$$p(a, b|x) = \frac{1}{2} q_1(a|x) r_1(b) + \frac{1}{2} q_2(a|x) r_2(b) \quad (31)$$

where $r_1(0) = 1, r_1(1) = 0, r_2(0) = 0, r_2(1) = 1$, and

$$q_1(a|x) = \begin{cases} 1 & \text{if } a = x = 0 \\ \frac{1}{N} & \text{if } a \geq 1, x = 1 \\ 0 & \text{else} \end{cases} \quad (32)$$

$$q_2(a|x) = \begin{cases} 1 & \text{if } a = 0, x = 1 \\ \frac{1}{N} & \text{if } a \geq 1, x = 0 \\ 0 & \text{else} \end{cases} \quad (33)$$

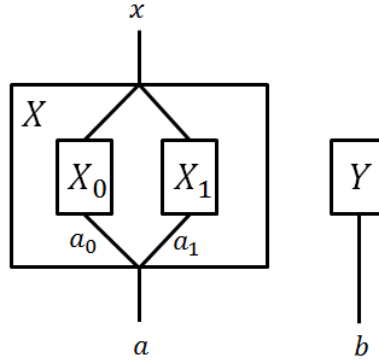
Consequently, the theorem could just as easily have read:

Theorem 7'. *Every bipartite GNST entropy vector is in the closure of the set of entropy vectors of separable GNST states.*

Suppose that we are given a bipartite GNST state and told its entropy vector, which is known to be accurate to within some ϵ , which can be arbitrarily small. Then the theorem tells us that we gain no knowledge of the non-local properties of the state. Whatever the entropy vector, the state may or may not be separable. This is in stark contrast with the von Neumann entropy, for which any vector with $S(AB) < S(A)$ instantly reveals that the state is entangled.

A further word of clarification. We showed that the measurement entropy in GNST satisfies only subadditivity and non-negativity and (at least for 2 parties) no further inequalities. (In fact we have evidence leading us to conjecture that this could also be true for 3 parties). This shows that entropy vectors in GNST are more general than entropy vectors in quantum theory. It would have been tempting to conclude that the reason for this is the extra non-locality available in GNST, but we have seen that this is not the case. What, then, is the cause?

Consider the following implementation of the state (27)-(28) using classical random variables.



In this system there are three classical random variables X_0, X_1 and Y . X_0 and X_1 are concealed within box X and are arranged such that:

- If Alice inputs $x = 0$, then $a = a_0$ and X_1 is destroyed.
- If Alice inputs $x = 1$, then $a = a_1$ and X_0 is destroyed.

The distribution of X_0, X_1, Y is as follows:

$$p(a_0, a_1, b) = \begin{cases} \frac{1}{2N} & \text{if } a_0 = 0, a_1 \geq 1, b = 0 \text{ or } a_0 \geq 1, a_1 = 0, b = 1 \\ 0 & \text{else} \end{cases} \quad (34)$$

which gives the same distribution as (27)-(28) for $p(a, b|x)$.

This raises an obvious question: if this GNST state can be realized via classical probability theory, why can't its entropies also be obtained there? The Shannon entropy vector of (34), considered as a bipartite state, is $(1 + \log N, 1, 1 + \log N)$ compared with the GNST entropy vector $(1 + \frac{1}{2} \log N, 1, 1)$. But the Shannon entropy is none other than the measurement entropy in the classical setting. Since the same measurement that achieved $\hat{H}(XY) = 1$ in GNST can also be performed classically, surely also $H(XY) \leq 1$?

The reason this is not true is that, although this measurement can be performed in the classical setting, it is not *maximally informative* there; since classically a maximally informative measurement must give the outputs of both X_0 and X_1 . This mechanism by which GNST artificially makes measurements on classical random variables which are not maximally informative to be so, by hiding some of the variables within the boxes, is the reason that GNST entropy vectors are more general than classical ones.

VI. DISCUSSION

When faced with the task of naming his entropy, Shannon was apparently told by von Neumann to call it 'entropy' because "nobody knows what entropy really is, so in a debate you

will always have the advantage”. Sixty years on this is still true. If ‘entropy’ means ‘the minimum amount of uncertainty of a system under a maximally informative measurement’ (and we accept the Shannon entropy as synonymous with ‘uncertainty’) then we are forced to accept the measurement entropy as the unique entropy in any physical theory. However, this is not the prevailing definition. If we take a more pragmatic approach, and allow any function to be deemed an ‘entropy’ if it satisfies a certain set of properties, then the question becomes: which properties do we choose?

This is where the result of the previous section fits. If you consider links between the von Neumann entropy and non-locality to be a happy coincidence, then this result has no bearing on the measurement entropy. If, however, you consider that in a highly non-local theory, such as GNST, an ‘entropy’ ought to reflect this non-locality, then the measurement entropy cannot really be an ‘entropy’.

It would be interesting to explore the existence of a ‘better’ entropy than measurement entropy in GNST. We know from [25] that there is no function which obeys all the same desirable properties that the von Neumann entropy does in the quantum regime. However, could there be a function with the same desirable properties as the measurement entropy that also detects non-locality? Another interesting problem would be to determine whether or not the analogue of Theorem 7’ holds in other GPTs which are between quantum theory and GNST.

Ultimately, the Shannon and von Neumann entropies are useful functions, not because of the desirable properties they have, but because of their impact on physics and information theory. Their importance lies in the fact they can be used to give expressions for optimal rates of compression, or for channel capacities. To the best of our knowledge, only one such theorem is known using the measurement entropy [17]. Other such theorems would be the best way to prove the usefulness of ‘entropy’ measures.

Acknowledgments

We thank Andreas Winter for very helpful discussions throughout the work, and especially Miguel Navascues who drew our attention to the separability of the states in section 4. The authors acknowledge support by the U.K. EPSRC and the European Commission (STREP project “QCS”).

-
- [1] C.E. Shannon. A mathematical theory of communication. *Bell System Tech. J.*, 27:379–423,623–656, 1948.
 - [2] J. von Neumann. Thermodynamik quantenmechanischer Gesamtheiten. *Nachr. Ges. Wiss. Göttingen, Math.-Phys. Klasse*, pages 273–291, 1927.
 - [3] C.H. Bennett, H.J. Bernstein, S. Popescu, and B. Schumacher. Concentrating partial entanglement by local operations. *Phys. Rev. A*, 53(4):2046–2052, 1996.
 - [4] S. Popescu and D. Rohrlich. Thermodynamics and the measure of entanglement. *Phys. Rev. A*, 56(5):R3319–R3321, 1997.
 - [5] M. Wilde. *From Classical to Quantum Shannon Theory*. arXiv:1106.1445, 2011.
 - [6] N. Pippenger. The inequalities of quantum information theory. *IEEE Trans. Inf. Theory*, 49(4):773–789, 2003.
 - [7] N. Linden and A. Winter. A new inequality for the von Neumann entropy. *Commun. Math. Phys.*, 259:129–138, 2005.
 - [8] J. Cadney, N. Linden, and A. Winter. Infinitely many constrained inequalities for the von Neumann entropy. *IEEE Trans. Inf. Theory*, 58(6):3657–3663, 2012.
 - [9] Z. Zhang and R.W. Yeung. On the characterization of entropy function via information inequalities. *IEEE Trans. Inf. Theory*, 44(4):1440–1452, 1998.

- [10] F. Matúš. Infinitely many information inequalities. In *IEEE Int. Symp. Inf. Theory*, pages 41–44, 2007.
- [11] R. Dougherty, C. Freiling, and K. Zeger. Non-Shannon information inequalities in four random variables. *arXiv:1104.3602*, 2011.
- [12] H. Araki and E.H. Lieb. Entropy inequalities. *Commun. Math. Phys.*, 18(2):160–170, 1970.
- [13] R.W. Yeung. *A First Course in Information Theory*. Kluwer.
- [14] Z. Zhang and R.W. Yeung. A non-Shannon-type conditional inequality of information quantities. *IEEE Trans. Inf. Theory*, 43(6):1982–1986, 1997.
- [15] B. Ibinson. *Quantum Information and Entropy*. PhD thesis, University of Bristol, 2007.
- [16] J. Barrett. Information processing in generalized probabilistic theories. *Phys. Rev. A*, 75:032304, 2007.
- [17] A.J. Short and S. Wehner. Entropy in general physical theories. *New J. Phys.*, 12:033023, 2010.
- [18] A.J. Short and J. Barrett. Strong nonlocality: A trade-off between states and measurements. *New J. Phys.*, 12:033034, 2010.
- [19] R. Horodecki, P. Horodecki, and M. Horodecki. Quantum α -entropy inequalities: independent condition for local realism? *Phys. Lett. A*, (210):377–381, 1996.
- [20] J.S. Bell. On the Einstein Podolsky Rosen paradox. *Physics*, 1:195–200, 1964.
- [21] S. Popescu and D. Rohrlich. Quantum nonlocality as an axiom. *Found. Phys.*, (24):379–385, 1994.
- [22] B.S. Tsirelson. Quantum generalizations of Bell’s inequality. *Lett. Math. Phys.*, (4):93–100, 1980.
- [23] H. Barnum, J. Barrett, L. Clark, M. Leifer, R. Spekkens, N. Stepanik, A. Wilce, and R. Wilke. Entropy and information causality in general probabilistic theories. *New J. Phys.*, 12:033024, 2010.
- [24] R.B. Ash. *Information Theory*. Dover, 1965.
- [25] S.W. Al-Safi and A.J. Short. Information causality from an entropic and a probabilistic perspective. *arXiv:1107.4031*, 2011.

Appendix A: Appendix

In this appendix we prove two of the lemmas stated in section III.

Lemma 2. *Let $A_1, A_2, B_1, \dots, B_n$ be a system of boxes, where only A_1 and A_2 are not classical. Then all measurements on this system are basic, or mixtures of basic measurements.*

Proof. Let $\mathbf{a} = (a_1, a_2)$ denote the outputs of the 2 non-classical boxes, and $\mathbf{x} = (x_1, x_2)$ denote their inputs. Let $\mathbf{b} = (b_1, \dots, b_n)$ denote the outputs of the classical boxes (since these boxes have no choice of input, we omit their inputs from the notation). Let $M = \{(r, \mu_r)\}$ be an arbitrary measurement on the system, with $\{\mathbf{R}_r\}$ a set of representing vectors given by Lemma 1.

For each \mathbf{b} let $\mathbf{R}_r^{(\mathbf{b})}$ be the vector with components $R_r^{(\mathbf{b})}(\mathbf{a}|\mathbf{x}) = R_r(\mathbf{a}, \mathbf{b}|\mathbf{x})$. Now, for fixed $\hat{\mathbf{b}}$, we claim that $\{\mathbf{R}_r^{(\hat{\mathbf{b}})}\}$ represent a measurement, $M^{(\hat{\mathbf{b}})}$ on the non-classical part of the system. To see this, note that whenever $p(\mathbf{a}|\mathbf{x})$ is a state on A_1, A_2 , we have:

$$\sum_{r, \mathbf{a}, \mathbf{x}} p(\mathbf{a}|\mathbf{x}) R_r^{(\hat{\mathbf{b}})}(\mathbf{a}|\mathbf{x}) = \sum_{r, \mathbf{a}, \mathbf{b}, \mathbf{x}} p(\mathbf{a}|\mathbf{x}) \delta_{\mathbf{b}\hat{\mathbf{b}}} R_r(\mathbf{a}, \mathbf{b}|\mathbf{x}) = 1 \quad (\text{A1})$$

where the last equality follows from the fact that $p(\mathbf{a}|\mathbf{x}) \delta_{\mathbf{b}\hat{\mathbf{b}}}$ is an allowed state of the overall system.

Since $M^{(\hat{\mathbf{b}})}$ is a measurement on a two box system, it must be a mixture of basic measurements [18]. The following is, therefore, also a mixture of basic measurements on $A_1, A_2, B_1, \dots, B_n$.

- (i) Obtain outputs \mathbf{b} .
- (ii) Perform measurement $M^{(\mathbf{b})}$.

But, in fact, this measurement is M , since for this measurement:

$$\text{Prob}(r) = \sum_{\mathbf{b}} p(\mathbf{b}) \text{Prob}(r|\mathbf{b}) \quad (\text{A2})$$

$$= \sum_{\mathbf{b}} p(\mathbf{b}) \sum_{\mathbf{a}, \mathbf{x}} R_r^{(\mathbf{b})}(\mathbf{a}|\mathbf{x}) p(\mathbf{a}|\mathbf{b}, \mathbf{x}) \quad (\text{A3})$$

$$= \sum_{\mathbf{b}} p(\mathbf{b}) \sum_{\mathbf{a}, \mathbf{x}} R_r^{(\mathbf{b})}(\mathbf{a}|\mathbf{x}) \frac{p(\mathbf{a}, \mathbf{b}|\mathbf{x})}{p(\mathbf{b})} \quad (\text{A4})$$

$$= \sum_{\mathbf{a}, \mathbf{b}, \mathbf{x}} R_r(\mathbf{a}, \mathbf{b}|\mathbf{x}) p(\mathbf{a}, \mathbf{b}|\mathbf{x}) \quad (\text{A5})$$

which is the same as the probability of getting outcome r in measurement M . Here, $p(\mathbf{a}|\mathbf{b}, \mathbf{x})$ is the probability of getting output \mathbf{a} from systems A_1, A_2 when we input \mathbf{x} given knowledge of the outputs \mathbf{b} from the classical boxes. This is equal to $p(\mathbf{b})^{-1} p(\mathbf{a}, \mathbf{b}|\mathbf{x})$ by the rules of conditional probability and the no-signalling condition. \square

Lemma 3. *A measurement $M = \{(r, \mu_r)\}$ is maximally informative if and only if every effect μ_r can be represented by a vector with only one non-zero entry, which is between 0 and 1.*

In order to make the proof of this lemma more clear, we first introduce two simple lemmas.

Lemma 8. *Let $(\mathbf{a}_1, \mathbf{x}_1)$ and $(\mathbf{a}_2, \mathbf{x}_2)$ be output-input pairs of a GNST system, with not both $\mathbf{a}_1 = \mathbf{a}_2$ and $\mathbf{x}_1 = \mathbf{x}_2$. Then there exists an allowed state, \mathbf{p} , such that $p(\mathbf{a}_1|\mathbf{x}_1) = 0$ and $p(\mathbf{a}_2|\mathbf{x}_2) > 0$.*

Proof. First suppose that $\mathbf{a}_1 \neq \mathbf{a}_2$. Then we can choose \mathbf{p} to be the distribution:

$$p(\mathbf{a}|\mathbf{x}) = \delta_{\mathbf{a}\mathbf{a}_2} \quad (\text{A6})$$

where $\delta_{\mathbf{x}\mathbf{y}}$ is 1 if $\mathbf{x} = \mathbf{y}$ and 0 otherwise.

Now suppose that $\mathbf{a}_1 = \mathbf{a}_2$. This means that we must have $\mathbf{x}_1 \neq \mathbf{x}_2$. Suppose that \mathbf{x}_1 and \mathbf{x}_2 disagree in the i th entry. Let $\bar{\mathbf{a}}_1$ be a vector of outputs which disagrees with \mathbf{a}_1 only in the i th entry. We can then choose \mathbf{p} to be the distribution:

$$p(\mathbf{a}|\mathbf{x}) = \begin{cases} \delta_{\mathbf{a}\bar{\mathbf{a}}_1} & \text{if } \mathbf{x}, \mathbf{x}_1 \text{ agree in } i\text{th entry} \\ \delta_{\mathbf{a}\mathbf{a}_1} & \text{else} \end{cases} \quad (\text{A7})$$

\square

Lemma 9. *Suppose that \mathbf{R} and \mathbf{S} are vectors representing the effect μ , such that \mathbf{R} has exactly one non-zero entry, and \mathbf{S} has no negative entries. Then $\mathbf{R} = \mathbf{S}$.*

Proof. Let $\mathbf{d} = \mathbf{R} - \mathbf{S}$. Since \mathbf{R}, \mathbf{S} both represent the same effect, it must be the case that $\mathbf{d} \cdot \mathbf{p} = 0$ for all states \mathbf{p} . We aim to show that $\mathbf{d} = \mathbf{0}$.

Let $(\mathbf{a}_1, \mathbf{x}_1)$ be such that $R(\mathbf{a}_1|\mathbf{x}_1) > 0$. Let $(\mathbf{a}_2, \mathbf{x}_2)$ be a distinct, but otherwise arbitrary, output-input pair. Note that $d(\mathbf{a}_2|\mathbf{x}_2) \leq 0$. Now choose \mathbf{p} according to the previous lemma, and notice that for this choice of \mathbf{p} , $\mathbf{d} \cdot \mathbf{p}$ will be negative, unless $d(\mathbf{a}_2|\mathbf{x}_2) = 0$. But $\mathbf{a}_2, \mathbf{x}_2$ were arbitrary, so in fact $d(\mathbf{a}_1|\mathbf{x}_1)$ is the only possibly non-zero component of \mathbf{d} . Finally, let \mathbf{p} be the distribution $p(\mathbf{a}|\mathbf{x}) = \delta_{\mathbf{a}\mathbf{a}_1}$ and then $\mathbf{d} \cdot \mathbf{p} = 0$ implies that, in fact, $\mathbf{d} = \mathbf{0}$. \square

Proof of lemma 3. Suppose that μ is an effect which can be represented by a vector, \mathbf{R} , with only one non-zero entry. Suppose also that $\mu = \sum_i \nu_i$ for some effects ν_i . For any vectors \mathbf{S}_i which represent ν_i , the vector $\sum_i \mathbf{S}_i$ represents μ . By lemma 1 we can choose the \mathbf{S}_i so that they have no negative entries. Then, by lemma 9, this means that $\mathbf{R} = \sum_i \mathbf{S}_i$. This implies that $\mathbf{S}_i \propto \mathbf{R}$, and hence $\nu_i \propto \mu$ for all i . Thus, if all the effects of a measurement can be represented in this way, then the measurement must be maximally informative.

Conversely, suppose μ cannot be represented by such a vector. Let \mathbf{R} be a vector with entries between 0 and 1 which represents μ . \mathbf{R} must have more than one non-zero entry. Let \mathbf{R}_1 be the vector which shares \mathbf{R} 's first non-zero entry, and has zeroes elsewhere, and let $\mathbf{R}_2 = \mathbf{R} - \mathbf{R}_1$. Then \mathbf{R}_1 and \mathbf{R}_2 both represent valid effects ν_1, ν_2 with $\nu_1 + \nu_2 = \mu$. Now, if $\nu_1 \propto \mu$ then for some constant λ we have $\lambda \nu_1 = \mu$ and hence $\lambda \mathbf{R}_1$ represents μ . But by lemma 9 this implies that $\lambda \mathbf{R}_1 = \mathbf{R}$, which is clearly false. Consequently, there must exist a non-trivial refinement of any measurement containing μ . \square

Remark. In the proof of lemma 8 (and hence also in lemmas 3 and 9) we assumed for simplicity that each box has more than one possible output. In the (rather trivial) case that some boxes have only one output, lemmas 8 and 9 do not hold. However, it is still possible to obtain lemma 3 by similar reasoning.

The key observation is the following. Suppose that we have a system of boxes, some of which have fixed outputs. We denote these boxes by X , their inputs \mathbf{x} and their outputs $\mathbf{0}$. The remainder of the system, Y , has boxes with inputs \mathbf{y} and outputs \mathbf{b} . We now show that we can reduce the theory to one on system Y only. The no signalling constraints ensure that for any state \mathbf{p} of the system, for all \mathbf{x}, \mathbf{x}' , $p(\mathbf{0}, \mathbf{b}|\mathbf{x}, \mathbf{y}) = p(\mathbf{0}, \mathbf{b}|\mathbf{x}', \mathbf{y})$. This implies that a vector \mathbf{R} represents an effect μ if and only if the vector \mathbf{R}' also represents μ , where

$$R'(\mathbf{0}, \mathbf{b}|\mathbf{x}, \mathbf{y}) = \begin{cases} \sum_{\mathbf{x}'} R(\mathbf{0}, \mathbf{b}|\mathbf{x}', \mathbf{y}) & \text{if } \mathbf{x} = \mathbf{0} \\ 0 & \text{else} \end{cases} \quad (\text{A8})$$

Thus the effect is essentially an effect on system Y : $R''(\mathbf{b}|\mathbf{y}) := R'(\mathbf{0}, \mathbf{b}|\mathbf{0}, \mathbf{y})$. We can now run the proofs of lemmas 3, 8 and 9 for this effect.